Large Language Model Routing with Benchmark Datasets

Year: 2023 | *Citations: 105* | *Authors: Tal Shnitzer, Anthony Ou, M'irian Silva*

Abstract

There is a rapidly growing number of open-source Large Language Models (LLMs) and benchmark datasets to compare them. While some models dominate these benchmarks, no single model typically achieves the best accuracy in all tasks and use cases. In this work, we address the challenge of selecting the best LLM out of a collection of models for new tasks. We propose a new formulation for the problem, in which benchmark datasets are repurposed to learn a"router"model for this LLM selection, and we show that this problem can be reduced to a collection of binary classification tasks. We demonstrate the utility and limitations of learning model routers from various benchmark datasets, where we consistently improve performance upon using any single model for all tasks.